心理科学进展 2022, Vol. 30, No. 10, 2143–2153
Advances in Psychological Science
https://doi.org/10.3724/SP.J.1042.2022.02143

·研究构想(Conceptual Framework)·

基于人类注意机制的微表情检测方法*

李婧婷[1] 东子朝[1] 刘 烨[1,2] 王甦菁[1,2] 庄东哲[3]

([1]中国科学院行为科学重点实验室(中国科学院心理研究所), 北京 100101)
([2]中国科学院大学心理学系, 北京 100039)
([3]中国人民公安大学公共安全行为科学实验室, 北京 100038)

摘 要 微表情是一种持续时间极短、不易被察觉的面部动作, 揭示了个体的真实情绪, 可以被广泛地应用于谎言识别等领域。而微表情检测的研究受到小样本问题的限制。针对该问题, 本文结合计算机视觉技术与认知心理学实验方法进行探索。首先, 结合眼动技术和呈现−判断范式与阈下情绪启动效应的行为实验范式, 考察微表情识别中选择注意分配的认知机制, 细化人类识别微表情时的特征兴趣区域。其次, 结合人类注意机制, 提出基于自监督学习的多模态微表情检测方法。通过理论和关键技术的突破, 为真实场景下微表情检测的应用奠定基础。

关键词 微表情检测, 小样本问题, 人类注意机制, 自监督学习, 深度信息
分类号 B842

1 问题提出

微表情是一种持续时间极短、不易被察觉的面部动作, 往往会在高风险情况下产生(Ekman & Friesen, 1969)。它由 Haggard 和 Isaacs 于 1966 年发现(Haggard & Isaacs, 1966), 之后由 Ekman 和 Friesen 命名。微表情的持续时间通常在 1/25 到 1/5 秒之间(Ekman & Friesen, 1969), 是非常重要的非语言交流线索, 它可以揭示真实的情绪和个人的心理状态(Haggard & Isaacs, 1966)。作为谎言识别的重要线索之一, 微表情的有效性甚至显著高于言语内容、语音、语调、身体姿势等其他线索(Owayjan et al., 2012), 可以被广泛地应用于国家安全、司法实践、临床诊断、学生教育、卫生防疫等领域。例如微表情可以作为重要线索来帮助此次新冠疫情的排查工作, 包括甄别人员是否对旅行史、密切接触情况以及发热症状等有所隐瞒。

持续时间短和动作幅度低是微表情的主要特征。如图 1 所示, 被试通过极其轻微的皱眉流露出厌恶的情绪, 但观察者单纯通过肉眼很难在视频中捕捉人脸微表情。微表情的人工检测和识别需要花费大量的人力物力, 并且参与分析的人员需要经过专业的训练。Ekman 在 2002 年开发了第一个微表情培训工具(Micro-Expression Training Tool, METT) (Ekman, 2003), 其目的是训练人类检测和识别微表情。然而即使是经过培训的专家, 其肉眼的识别率也低于 50% (Ekman, 2003)。因此, 为了将微表情分析应用到实际场景中, 对微表情智能检测与识别的研究十分迫切。

本研究将通过计算机和心理学的交叉, 研发人脸微表情智能检测系统, 具有重要的理论意义和巨大的应用价值。针对微表情小样本问题, 首先在心理学领域, 研究面向微表情的注意力机制, 高效提取微表情特征, 提升系统对微表情动作的敏感度和可靠性。其次, 在计算机视觉技术领域, 针对小样本问题, 除了进一步扩充数据容量, 在

收稿日期: 2022-03-24
* 国家自然科学基金项目(62106256、U19B2032)、中国人民公安大学公共安全行为科学实验室开放课题(2020SYS12)和中国博士后科学基金项目(2020M680738)资助。
通信作者: 王甦菁, E-mail: wangsujing@psych.ac.cn;
庄东哲, E-mail: zdzfrued@126.com

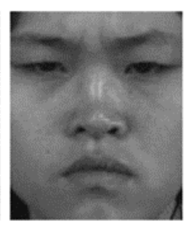


图 1 微表情样例：紧张情绪，眉心处轻微向内收拢，持续时间约 300 ms。(资料来源：CASME II 数据库, ©Xiaolan Fu)

算法层面，主要是通过将其余相似类型的数据特征进行迁移，或者通过挖掘数据自身的特征信息进行学习(无监督/自监督学习)。基于上述研究方法，本研究提出基于人类注意机制的多分支自监督学习模型，实现微表情智能检测，进一步避免了小样本问题对微表情研究的限制。此外，为了增强网络构建自监督信息的能力，本研究利用包含场景深度信息的微表情样本，实现多模态的微表情分析。

2 微表情智能分析研究现状

2.1 基于计算机视觉技术的微表情国内外研究现状及发展动态分析

近 10 年来，微表情研究越来越受到科学领域和大众媒体的重视，如图 2 所示，研究人员试图通过计算机科学与心理学的结合，研发自动检测和识别微表情的技术，以帮助人们更有效地利用微表情线索来识别谎言。微表情识别是对已知存在微表情的片段进行情感分类，技术相对成熟(See et al., 2019)。反之，微表情检测是在长视频中准确定位微小短暂的微表情片段，任务十分具有挑战性，并且相关的论文数量仍然相对较少，但其研究具有很高的实用价值。如果能在一段视频中准确地检测到某个时间点有微表情出现，那么就说明这个人在这个时刻可能会有异常。

由于微表情具有强度低和持续时间短的特点，微表情相关特征很难提取，这对微表情检测造成了一定的困难。同时，当前微表情检测算法的准确率不够高，即使是对严格控制的实验环境中采集的微表情视频进行检测，微小的面部运动也会导致误检。另外，因为微表情样本的人工标注十分困难，目前只有 6 个公开的自发微表情数据库，包括中国科学院心理研究所发布的 CASME 系列(Qu et al., 2018; Yan et al., 2013, 2014)，芬兰奥卢大学发布的 SMIC (Li et al., 2017)、英国曼彻斯特城市大学发布的 SAMM (Davison, Lansley, et al., 2018)和中国山东大学发布的 MMEW (Ben et al., 2021)，视频总量约 1000 个，这是典型的小样本问题。大数据驱动的深度学习已经在诸多领域获得成功，但是微表情小样本问题限制了基于深度学习的微表情检测研究。

因此，本研究面向微表情检测方法，针对微表情小样本问题，分别从注意力机制、基于大量人脸和表情数据库的自监督学习和深度信息三方面来解决这一问题。本节将首先介绍微表情检测的相关方法的研究现状，然后对本研究中应用的模式识别技术理论和方法(即注意力机制、自监督学习和深度信息)的国内外研究现状进行综述。

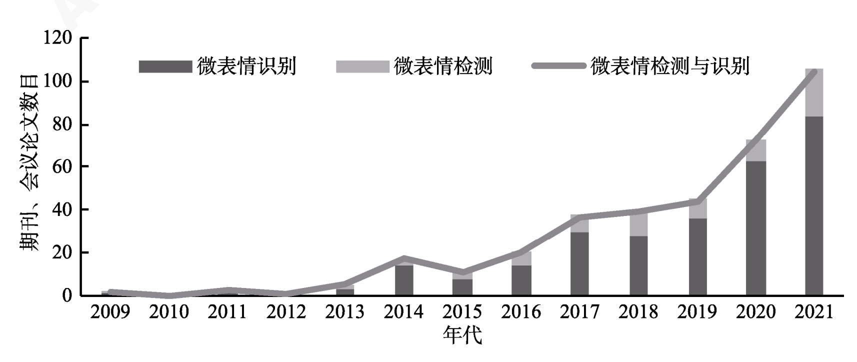


图 2 基于计算机的微表情研究趋势

2.1.1　微表情检测的研究现状

目前通过计算机技术对微表情进行分析成为了一个研究热点，国内很多团队都开展了相关的科研工作，比如中国科学院心理研究所傅小兰团队，清华大学刘永进团队，中国科学院自动化研究所陶建华团队，复旦大学张军平团队，上海交通大学李生红团队，东南大学郑文明、宗源团队，西安交通大学洪晓鹏团队，山东大学贲晛烨团队，西北工业大学冯晓毅、夏召强团队，北京交通大学安高云团队，江西中医药大学申寻兵团队，西南大学陈通团队，北京科技大学马惠敏团队、谢仑团队和支瑞聪团队，南京工程学院黄晓华团队，中国科学技术大学陈恩红团队，江苏大学毛启容团队，北京师范大学孙波团队，合肥工业大学詹曙团队，南京晓庄学院郑豪团队，江西师范大学鄢克雨团队等。然而，大部分团队主要研究了微表情识别方法，微表情检测相关研究仍有待探索。

微表情检测方法主要有两种思路，一种是通过比较帧间特征差异检测微表情，另一种是通过机器学习提取微表情的特征进而对微表情帧和非微表情帧进行分类。

目前大多数方法利用特征差异来检测微表情，主要流程是计算时间窗口中所提取特征的差异，通过在整个视频中设置阈值，可以发现最明显的脸部运动。常用的特征包括：芬兰奥卢大学赵国英团队使用的局部二值模式(LBP) (Moilanen et al., 2014)，马来西亚多媒体大学的梁诗婷等人(Liong et al., 2015)、以及王甦菁团队(Wang et al., 2016)使用的光流相关特征，英国曼彻斯特城市大学 Moi Hoon Yap 团队使用的 3D-HOG (Davison, Merghani, et al., 2018)等。这些方法的主要优点是能够在基于微表情持续时间的滑动窗口内进行帧之间的比较。然而，通常情况下，仅窗口中的第一帧和最后一帧被用于当前帧的特征差计算，这种方法没有考虑到微表情的时间变化模式。特征差异方法的另一个缺点是无法区分微表情和其他类型的头部运动，尤其是在长视频中，特征差异方法会发现许多高于阈值的运动，从而导致许多假阳性。

如今，为了增强检测方法区分微表情与其他面部运动的能力，基于机器学习/深度学习的微表情检测方法刚刚兴起。目前只有十余篇相关论文发表，例如 Wang 等人(2021)发表的 MESNet、北京科技大学谢仑团队提出的局部双线卷积神经网络(Pan et al., 2020)、中国电子科技集团电子科学研究院谢海永团队提出的基于光流和 LSTM 的检测方法(Ding et al., 2019)、荷兰埃因霍芬理工大学 Vlado Menkovski 团队提出的结合光流和 RNN 的检测方法(Verburg & Menkovski, 2019)、芬兰奥卢大学赵国英团队提出的 LBP-TOP+SVM 检测方法(Tran et al., 2017)等、Li 等人(Li, Soladie, et al., 2019; Li et al., 2020)提出具备微表情普适化的局部时域模式，进而通过机器学习分类器检测微表情。当前基于深度学习的微表情检测方法研究受到小样本问题限制，无法应用到实际场景中。

2.1.2　相关模式识别中的技术理论和方法研究现状

基于物体的注意力的神经机制，麻省理工学院 Baldauf 和 Desimone 于 2014 年在《Science》杂志上发表了其研究成果。注意力属于认知过程的一部分，是一种导致局部刺激的意识水平提高的知觉选择性的集中，是心理学中研究最热门的主题之一。计算机领域中注意力机制的提出使得深度网络能够忽略无关信息而关注和捕捉重点特征。在图像处理领域，相关研究主要分为强注意力(hard attention)和软注意力(soft attention)两种类型，其中强注意力关注每个点可能具备的注意力，并且强调图像的动态变化，目前主要应用在图像裁剪领域(Mnih et al., 2014)；软注意力则更关注区域或者通道，并且通过网络生成确定的注意力模块，该机制在深度网络上得到了广泛的应用(Fu et al., 2019; Li et al., 2018)。由于微表情微小且短暂的特性，Wen 等(2021)、Li 等(2021)以及其他研究团队尝试引入注意力机制，提升网络针对微表情特征提取的能力。然而，由于针对微表情自身的认知机制中注意力分配的研究较少，微表情特征提取的有效性需要进一步提升。

2015 年，LeCun、Bengio 和 Hinton 联合在《Nature》杂志发表了一篇深度学习的综述文章(LeCun et al., 2015)。文章最后指出，像人类视觉系统一样的无监督的深度学习是一个重要的方向。自监督学习是无监督学习的一种方法(Jing & Tian, 2020)，通过在大规模的无监督数据中利用辅助任务挖掘自身的监督信息，并利用这种构造的监督信息对网络进行训练，从而可以学习到对下游任务有价值的特征信息。尽管单个的无标注

样本相较于已标注的样本所包括的信息要少，但是如果能获取大量无监督数据，并且算法可以有效地通过辅助任务获取相应的监督信息，那么相比大规模的耗时耗力的手工特征构建和数据标注，自监督算法将有更好的效率和实用性。自监督学习已经在图片 Patch 相对位置预测(Doersch et al., 2015)、图像修复(Pathak et al., 2016)、图片旋转角度预测(Gidaris et al., 2018)、图像色彩预测(Larsson et al., 2017)、视频中目标检测(Li, Liu, et al., 2019)、视频顺序预测(Fernando et al., 2017)等任务中实现利用无监督数据构造自身监督信息，获得了可以和监督学习性能媲美的结果。

深度信息为图像分析提供了更多的几何信息，有助于提升人脸与表情识别的可靠性。国内外对其展开了大量的研究，例如 Danelakis 等(2016)提出了利用面部坐标的 GeoTopo+特征，以创建三个子特征来捕获面部的拓扑和几何信息；国内厦门大学雷蕴奇团队提出了结合深度图和 SURF 算法的 3D 人脸识别(Ma et al., 2017)；中国科学技术大学於俊团队提出了基于深度图的姿势估计和人脸重构(Luo et al., 2019)等。由于面部肌肉运动会引起其对应深度信息的变化，引入深度信息可以帮助系统更加灵敏地检测到人脸表情的变化。通过借鉴结合深度图的人脸表情研究(Cai et al., 2020)，将有助于开展结合深度信息的人脸微表情检测的分析。

2.2 本文贡献

由于微表情是一种持续时间极短、不易被察觉的面部动作，导致了微表情的诱发、采集和人工标注十分困难。样本采集和标注往往需要大量的人力物力，成本很高。这一情况限制了微表情数据库的创建，该小样本问题桎梏了结合机器学习的微表情检测方法的研究。如图 3 所示，本研究通过研究注意力机制增强网络针对微表情特征提取能力，以及通过自监督学习的方法实现微表情检测，从而避免了小样本问题的限制。此外，由于微表情的特性，现实生活中的很多视频中存在大量未标注的微表情，例如在访谈节目中，当被访问者不想透露真实情绪时，往往脸部会出现微表情。本研究提出的方法能够通过在无标注视频中挖掘微表情片段，实现微表情样本量的扩充，从而解决微表情小样本问题，并促进传统微表情检测方法的性能提升。

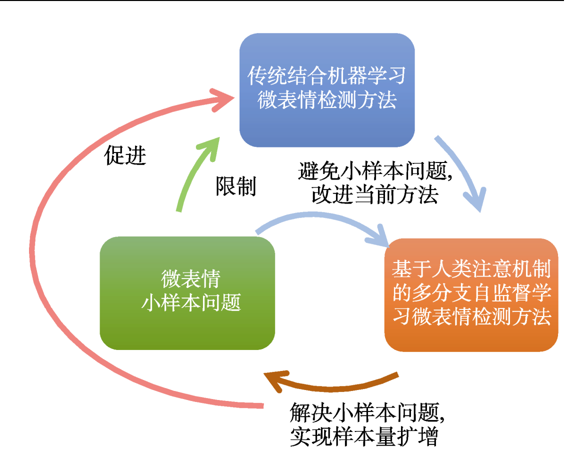


图 3 本研究提出的微表情检测方法与微表情小样本问题关系图

首先，本研究开展基于心理学微表情认知机制中注意力资源的相关探索，发现针对微表情特性的注意力机制，从而有效、直接地使模型从输入信息中就获取微表情重要特征，并通过网络从多维度(时域、空间域和通道域)进一步生成相关注意力模块，在样本量有限的情况下，提升网络对微表情特征的提取性能。

其次，本研究提出基于人类注意机制的多分支自监督学习方法，在预训练的过程中从大量无标注视频中学习时空特征，从而实现下游任务即微表情的检测。这一方法避免了小样本问题对传统有监督机器学习方法的限制，并且该方法具有普适性，为后续不同实际场景的微表情检测提供了技术基础。

最后，目前的微表情发布数据均为视频样本，并没有包括相应的深度信息。本研究将基于我们研究团队正在创建的首个包含图像深度信息的微表情数据库，开展基于深度信息的微表情检测，为自监督学习微表情检测模型的辅助任务提供更多信息。

本研究提出的基于人类注意机制的多分支自监督学习的微表情检测方法，避免了小样本问题的限制，使在国安审讯、医疗问诊等复杂真实场景下微表情分析技术的应用成为可能。

3 研究构想

3.1 基础理论和模型的研究构想

微表情检测在谎言识别、医疗诊断等领域都有着广阔的应用前景。然而，由于微表情样本采

集和标注困难，造成小样本问题，限制了微表情检测方法的研究。针对这一问题，本研究从两个方面开展研究，首先通过人类注意机制的研究，提升深度网络对微表情特征提取能力；其次，利用无标注样本训练自监督模型，构建普适且高效的微表情检测网络。图 4 展示了本研究的研究框架。

3.1.1　微表情识别的人类注意机制分析

(1)心理学认知微表情的注意力资源分配研究

首先，在心理学层面，面孔中的许多信息都会引导注意朝向。基于认知心理学的行为实验方法，采用呈现−判断范式，考察人在进行微表情识别时的注视线索。实验结论有助于细化微表情的特征兴趣区域，同时帮助解决微表情局部信息全局化的合理性。其次，阈下情绪启动效应是情绪启动效应的一种经典范式。利用微表情的表现方式与阈下情绪启动方式一致这一特点，将微表情识别分为无意识情绪启动阶段和情绪知觉理解阶段，并进行实验验证。前期无意识阶段的研究可以提升系统局部特征提取的能力，后期决策的研究可以帮助解决微表情特征全局空间整合的问题。

(2)针对微表情检测的注意力机制设计

在计算机视觉层面，利用针对微表情的注意朝向和大脑意识的认知阶段，帮助网络从输入信息中选择重要的特征，进而结合多维度的注意力机制，从时域、空间域以及通道域构建注意力模块，设计针对微表情局部特征提取和全局决策的注意力机制，提升局部信息整体融合的有效性，从算法层面优化受限于小样本问题的微表情检测方法。

3.1.2　多分支自监督学习的微表情检测研究

针对微表情训练样本少的问题，我们提出通过构建多分支自监督学习模型，设计辅助任务，使模型在大量包含人脸及发生面部表情的视频中构建自监督信息，提取微表情关键区域的时空特征，最后通过微调将模型应用到下游任务中。这种无监督学习模型的设计可以避免小样本问题对微表情检测的限制，并为适应后续实际场景的微表情检测提供可能。

除了传统的彩色图像(RGB)模态，深度信息是体现视点和场景对象表面之间距离的多模态信息，直接反应了可见表面的几何形状。由于微表情是脸部的微小运动，深度信息的补充可以帮助提取微表情特征。本研究将结合深度信息进行微表情检测方法的探索，通过利用深度流和 RGB 图像，仿真生成深度图来构建辅助任务，提升自监督学习提取监督信息的能力，从而增强算法检测微表情的性能。

3.1.3　基于人类注意机制的多分支自监督学习网络的微表情检测

通过结合基于人类注意机制和多分支自监督模型，本研究创建了能够高效提取针对微表情特征的无监督微表情检测方法，克服微表情小样本问题的限制，并有助于无标注视频中微表情样本的检测与时间标注(起始/终止帧)，从而实现微表情的样本扩增。此外，该自监督模型的建立为复杂场景下微表情检测的应用奠定了基础。

3.2　关键技术的研究构想

本研究融合了计算机视觉和认知心理学，属于模式识别系统及应用与认知心理学的跨学科交叉领域的研究。如图 5 所示，为了研究在小样本问题限制下的微表情检测方法，本研究通过基于心理学认知的注意力机制研究，优化微表情特征提取过程，并搭建基于人类注意机制的多分支自监督学习网络，从而实现无监督微表情检测。

3.2.1　微表情认知的注意力机制分析

在注意力机制研究部分，我们首先开展微表情识别的认知研究，探索认知过程中的注意力资源变化。然后，基于心理学原理和相关实验结果，

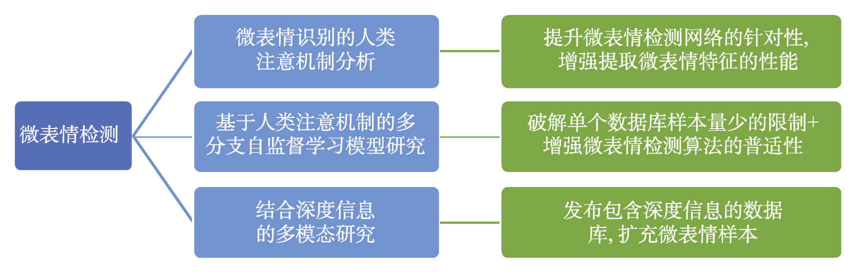


图 4　基于小样本问题的微表情检测研究框架

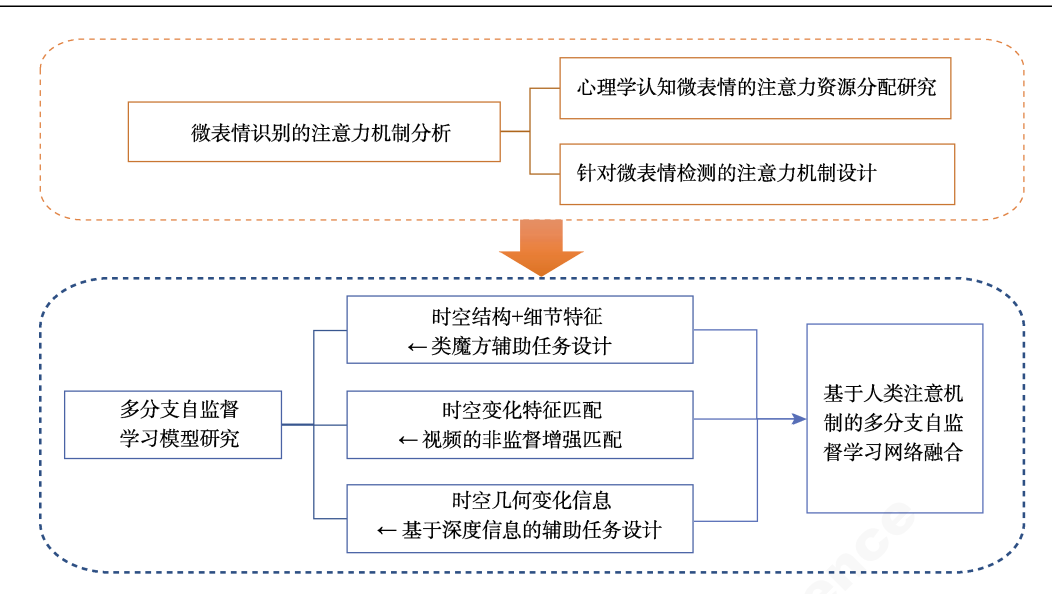


图 5 基于人类注意机制的多分支自监督学习微表情检测之技术框架

结合计算机视觉技术，在样本量有限的情况下，搭建针对微表情的注意力模块，提升计算模型学习和整合微表情特征的能力。

(1)心理学认知微表情的注意力资源分配研究

观察者的注意力朝向受到面孔中的诸多信息的影响。脑成像研究表明，与中性面孔相比，情绪面孔前注意阶段的加工可以导致视皮层激活增强，反映了对情绪面孔注意朝向的增强。因此，本研究基于认知心理学的行为实验方法，采用呈现−判断范式，考察人在进行微表情识别时的注视线索。同时，通过眼动仪记录被试识别微表情的视觉扫描特征。依据分析结果和 AU 单元，划分面孔刺激材料的兴趣区域，判断微表情识别时重要的提取线索区域，并进一步分析该区域的视觉扫描特征。

在微表情识别的加工过程中，不同阶段的注意朝向不同。我们通过借鉴阈下情绪启动方式的研究来分析微表情识别的注意力变化机制。在研究过程中，通过实验验证，微表情识别被分为两个阶段，分别为无意识情绪启动阶段和情绪知觉理解阶段。采用 ERP 实验，探讨在情绪不一致时微表情识别的脑加工机制。

通过以上的心理学理论研究和实验，我们可以获得有效且细化的微表情认知过程中兴趣区域变化过程，以及大脑注意力机制的启动和决策过程，从而帮助微表情检测系统实现全局→局部→全局的针对微表情注意力机制设计。

(2)针对微表情检测的注意力机制设计

由于微表情特征是局部分布在人脸上的，同时具备强度低和持续时间短的特点，引入注意力机制将有助于系统对微表情特征的提取。特别是，不同的微表情对应的重点脸部区域可能会不同。例如，已有研究表明对于高兴的面部表情，嘴部区域的动作更加重要；而对于恐惧的表情，眼睛区域的动作更加明显(Eisenbarth & Alpers, 2011)。针对微表情的注意力机制模块设计如图 6 所示，根据心理学的微表情认知机制，从输入信息中得到重点区域，并由通道域的注意力机制进行重点关注，随后依次通过时空注意力模块，提取微表情的有效特征，再依据心理学的决策机制，最终得到合理的融合性全局特征。此外，通过引入注意机制和无注意机制的对比实验，验证微表情识别注意研究的可靠性。

3.2.2 多分支自监督学习模型的微表情检测研究

由于已标注的微表情样本有限，本研究提出在大量的无标注人脸及表情视频中进行自监督学习，通过构建针对微表情特征的辅助任务，实现在下游的微表情检测任务。微表情是一个短暂的局部面部动作，除了空间特征，其时域变化特征对微表情检测也很重要。本研究通过构建多分支自监督学习模型，提取视频样本的时空特征，包括结构信息、细节信息、动态变化模式和深度几

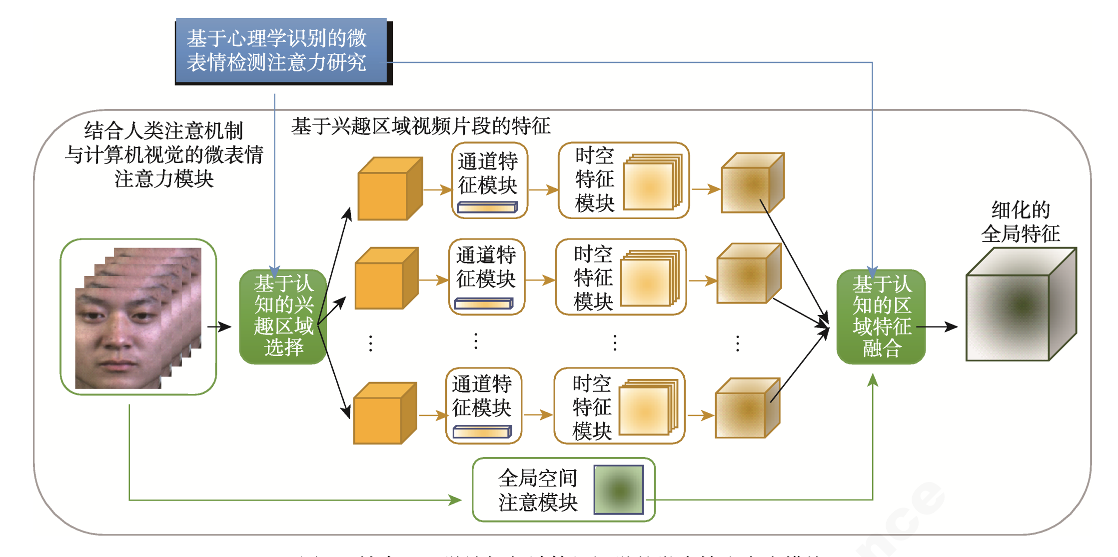


图 6　结合心理学认知与计算机视觉的微表情注意力模块

何信息，从而实现无监督的微表情检测。

1)受到心理学人类注意机制的启发，针对人脸微表情局部分布的特征，本研究设计了基于兴趣区域视频的排序及补全的类魔方辅助任务，分别在视频中提取对应心理学认知全局决策的结构信息和对应局部注意的细节信息。该类魔方辅助任务可以构建具备提取视频结构和细节特征能力的自监督网络，并且优化微表情检测方法的全局–局部策略。如图 7 所示，首先将常见发生微表情的局部区域从人脸图像中分离出来，通过随机排序规则，使模型在大量的样本训练中学习图像的结构信息。此外，为了学习人脸动作在时间上的变化，本文将人脸视频按照时间窗口进行分割，和空间的人脸局部区块结合，实现时空结构特征的提取。其次，在获取人脸视频结构信息的同时，通过补全视频片段辅助任务来获取人脸视频的细节信息。我们通过随机裁减掉一个兴趣区域，在网络学习局部区域视频结构特征的同时，利用自编码器网络对缺失视频的重构和补全来学习对应区域的细节信息。

2)相同类型的表情，由于个体差异(人种、年龄、性别等)、面部肌肉运动单元强度不同等因素，存在不同的表现形式。受启发于一维信号动态时间规整(Dynamic Time Warping, DTW)以及图像的非监督增强匹配技术在小样本问题中的应用，本研究提出基于视频的非监督增强匹配网络设计，借助时空结构保持的特性，通过已知表情和新出现表情视频样本的对比，实现跨表现形式的特征匹配。这样的时空结构匹配使得网络具备提取视频时空动态变化模式的能力，并且与心理学微表情认知局部→全局匹配的过程相呼应。

3)基于深度相机采集的深度图，我们获取了针对微表情视频样本的深度信息。连续帧在同一区域的深度值变化体现了当前区域的人脸几何变化信息。并且基于心理学发现，人的深度视觉是人对物体认知的重要参考因素。同时，在基于人类注意机制的人脸兴趣区域中，一方面，深度信息将有区别于其他区域的更明显的变化，一方面，检测系统本身也会通过集中关注这些区域从而获得更有效的微表情动作信息。因此，本研究利用深度信息设计自监督学习辅助任务，能够提升模型捕捉运动信息的能力。

基于上述三种辅助任务，本研究搭建了一个基于人类注意机制的多分支自监督学习模型。通过在大量无标注视频样本进行三个辅助任务的训练，使得模型具备在人脸视频中提取兴趣区域的结构特征+细节特征(类魔方任务)、视频动态变化模式(非监督增强匹配)和几何信息(深度信息估计)的能力。由此，如图 8 所示，本研究得到能够提取时空特征的多分支自监督学习模型，结合针对基于微表情认知的注意力模块，通过微调和结合全局注意力机制的特征细化，实现下游任务的微表情检测。

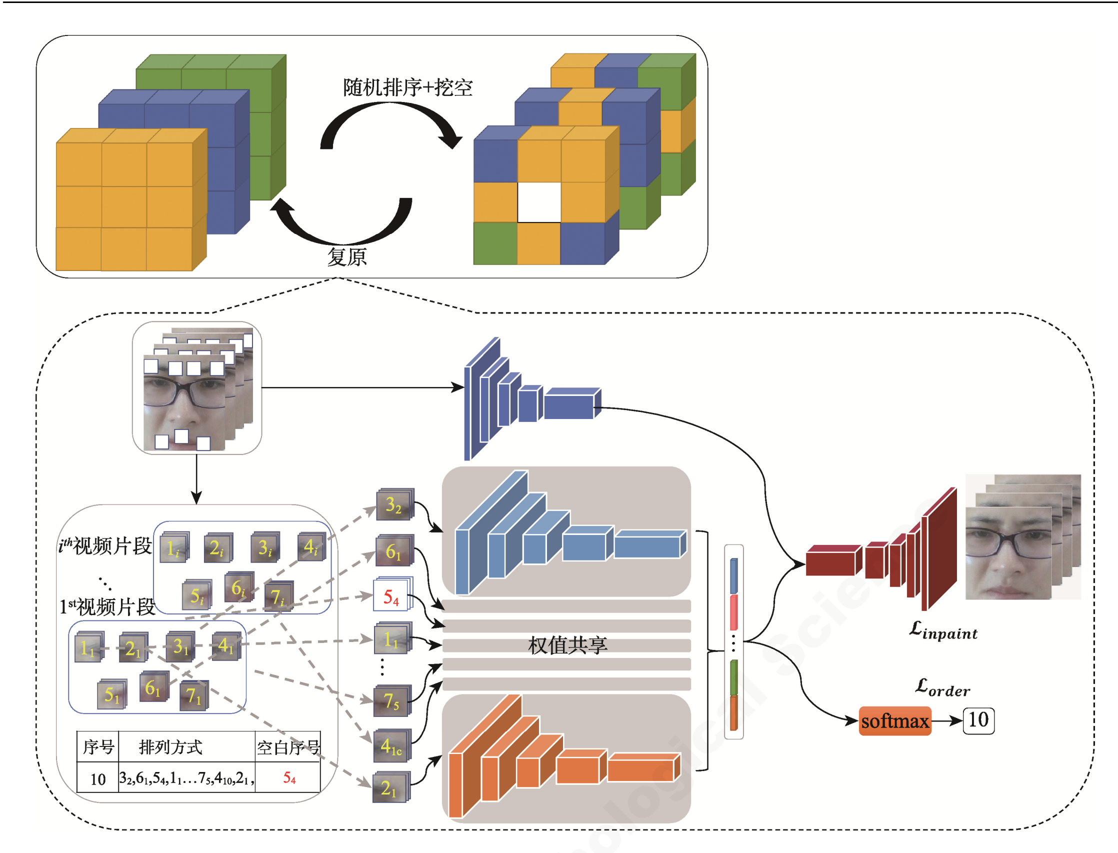


图 7 基于类魔方辅助任务的自监督网络框架

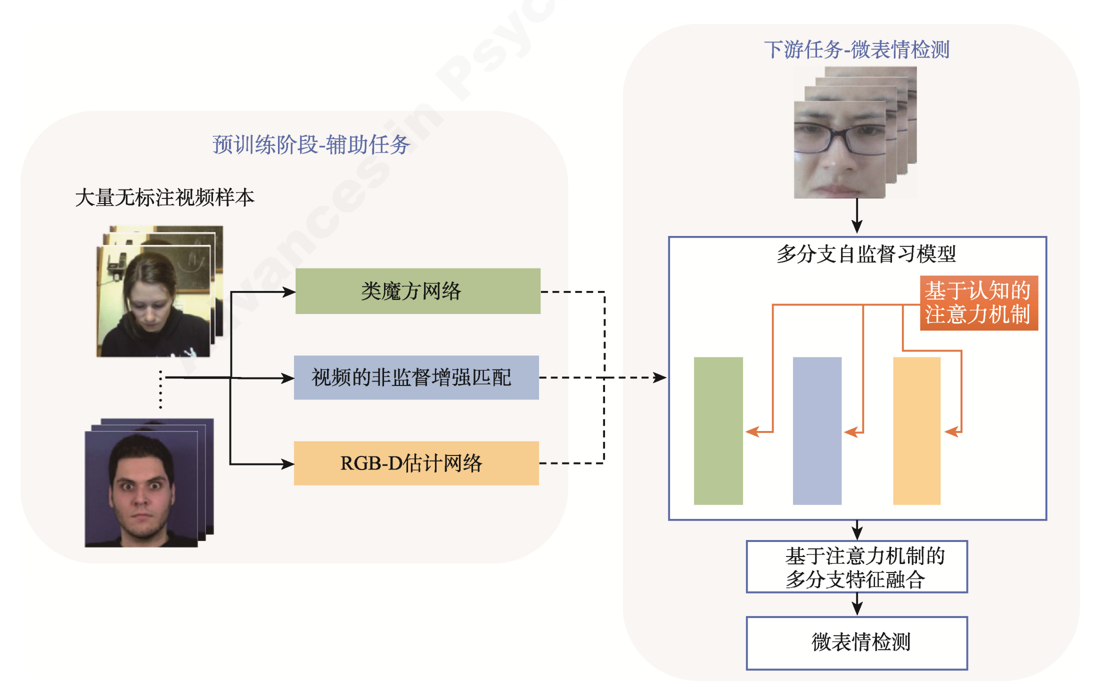


图 8 结合人类注意机制的多分支自监督学习的微表情检测

本研究提出的自监督学习模型通过三个并行的分支辅助任务，从平面结构、平面细节、时空动态变化和运动集合信息的四个维度，全面地提取了样本时空特征，从而实现下游任务的微表情检测。同时，在注意力机制部分，每个自监督网络的分支通过结合本研究提出的人类注意机制模块，增强了微表情的特征提取性能和全局特征融合的能力，并在最后通过注意力机制融合并细化的多分支特征，得到最具表征性的微表情特征。

本研究提出的基于人类注意机制的多分支自监督学习模型，克服了小样本问题对微表情研究的限制，并为后续适应不同场景的微表情检测提供可能。

4 理论构建与创新

微表情检测在深度学习领域的发展受到了小样本问题的限制。本文将计算机视觉技术与认知心理学行为实验方法相结合，自主研发具有重要科学研究和实际应用价值的微表情智能检测系统。

(a)基于微表情认知过程，通过对其注意力资源的研究，探索了针对微表情局部短暂的时空特征的多维度注意力机制，优化了网络特征提取的能力，在受限于微表情样本容量小的情况下，增强了微表情检测系统的敏感度和可靠性。

(b)首次将自监督学习方法引入到微表情检测模型构建中，通过在大量无监督视频中利用辅助任务构建具备提取微表情时空特征的模型，实现下游任务−微表情的检测，从而避免了标注样本量小带来的限制。此外，无监督微表情检测系统的实现能够在日常视频中挖掘微表情样本，扩充数据量，帮助解决小样本问题。

(c) 首次结合深度信息对微表情分析进行研究，并构建结合深度图的自监督学习模型，开展多模态微表情检测研究。

参考文献

Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, *344*(6182), 424–427.

Ben, X., Ren, Y., Zhang, J., Wang, S.-J., Kpalma, K., Meng, W., & Liu, Y.-J. (2021). Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2021.3067464

Cai, J., Xie, H., Li, J., & Li, S. (2020). Facial expression recognition with an attention network using a single depth image. In H. Yang, K. Pasupa, A. C.-S. Leung, J. T. Kwok, J. H. Chan, & I. King (Eds.), *Neural Information Processing* (pp. 222–231). Springer International Publishing.

Danelakis, A., Theoharis, T., Pratikakis, I., & Perakis, P. (2016). An effective methodology for dynamic 3D facial expression retrieval. *Pattern Recognition*, *52*, 174–185.

Davison, A., Merghani, W., Lansley, C., Ng, C. C., & Yap, M. H. (2018, May). Objective micro-facial movement detection using facs-based regions and baseline evaluation. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)* (pp. 642−649). IEEE.

Davison, A. K., Lansley, C., Costen, N., Tan, K., & Yap, M. H. (2018). SAMM: A spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing*, *9*(1), 116–129. https://doi.org/10.1109/TAFFC.2016.2573832

Ding, J., Tian, Z., Lyu, X., Wang, Q., Zou, B., & Xie, H. (2019, September). Real-time micro-expression detection in unlabeled long videos using optical flow and LSTM neural network. In *International Conference on Computer Analysis of Images and Patterns* (pp. 622−634). Springer-Verlag.

Doersch, C., Gupta, A., & Efros, A. A. (2015, December). Unsupervised visual representation learning by context prediction. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)* (pp. 1422−1430). IEEE.

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion*, *11*(4), 860–865. https://doi.org/10.1037/a0022758

Ekman, P. (2003). *Emotions revealed*. St. Martin's Griffin, New York.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, 32(1), 88–106.

Fernando, B., Bilen, H., Gavves, E., & Gould, S. (2017, July). Self-supervised video representation learning with odd-one-out networks. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3636−3645). IEEE.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., & Lu, H. (2019, June). Dual attention network for scene segmentation. In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3146−3154). IEEE.

Gidaris, S., Singh, P., & Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. *ArXiv Preprint ArXiv:1803.07728*.

Haggard, E. A., & Isaacs, K. S. (1966). Micromomentary

facial expressions as indicators of ego mechanisms in psychotherapy. In *Methods of research in psychotherapy* (pp. 154−165). Springer, Boston, MA.

Jing, L., & Tian, Y. (2020). Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(11), 4037–4058.

Larsson, G., Maire, M., & Shakhnarovich, G. (2017). Colorization as a proxy task for visual understanding. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6874−6883). IEEE.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Li, H., Xiong, P., An, J., & Wang, L. (2018). Pyramid attention network for semantic segmentation. *ArXiv Preprint ArXiv:1805.10180*.

Li, J., Soladie, C., & Seguier, R. (2020). Local temporal pattern and data augmentation for micro-expression spotting. *IEEE Transactions on Affective Computing*, pp.1–1. https://doi.org/10.1109/TAFFC.2020.3023821

Li, J., Soladie, C., Seguier, R., Wang, S. J., & Yap, M. H. (2019, May). Spotting micro-expressions on long videos sequences. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)* (pp. 1–5). IEEE.

Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & Pietikäinen, M. (2017). Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Transactions on Affective Computing, 9*(4), 563–577.

Li, X., Liu, S., de Mello, S., Wang, X., Kautz, J., & Yang, M.-H. (2019). Joint-task self-supervised learning for temporal correspondence. *Advances in Neural Information Processing Systems*, *32*.

Li, Y., Huang, X., & Zhao, G. (2021). Micro-expression action unit detection with spatial and channel attention. *Neurocomputing*, *436*, 221–231.

Liong, S.-T., See, J., Wong, K., Le Ngo, A. C., Oh, Y. H., & Phan, R. (2015, November). Automatic apex frame spotting in micro-expression database. In *2015 3rd IAPR Asian conference on pattern recognition (ACPR)* (pp. 665−669). IEEE.

Luo, C., Zhang, J., Yu, J., Chen, C. W., & Wang, S. (2019). Real-time head pose estimation and face modeling from a depth image. *IEEE Transactions on Multimedia*, *21*(10), 2473–2481.

Ma, J., Zhang, H., & She, W. (2017, June). Research on robust face recognition based on depth image sets. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)* (pp. 223−227). IEEE.

Mnih, V., Heess, N., & Graves, A. (2014). Recurrent models of visual attention. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 27). Curran Associates, Inc.

Moilanen, A., Zhao, G., & Pietikäinen, M. (2014, August). Spotting rapid facial movements from videos using appearance-based feature difference analysis. In *2014 22nd international conference on pattern recognition* (pp. 1722−1727). IEEE.

Owayjan, M., Kashour, A., Al Haddad, N., Fadel, M., & Al Souki, G. (2012, December). The design and development of a lie detection system using facial micro-expressions. In *2012 2nd international conference on advances in computational tools for engineering applications (ACTEA)* (pp. 33−38). IEEE.

Pan, H., Xie, L., & Wang, Z. (2020, November). Local bilinear convolutional neural network for spotting macro- and micro-expression intervals in long video sequences. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)* (pp. 749–753). IEEE.

Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016, June). Context encoders: Feature learning by inpainting. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2536−2544). IEEE.

Qu, F., Wang, S.-J., Yan, W.-J., Li, H., Wu, S., & Fu, X. (2018). CAS $(ME)^2$: A database for spontaneous macro-expression and micro-expression spotting and recognition. *IEEE Transactions on Affective Computing*, *9*(4), 424–436.

See, J., Yap, M. H., Li, J., Hong, X., & Wang, S. J. (2019, May). MEGC 2019–the second facial micro-expressions grand challenge. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)* (pp. 1–5). IEEE.

Tran, T.-K., Hong, X., & Zhao, G. (2017). Sliding window based micro-expression spotting: A benchmark. In J. Blanc-Talon, R. Penne, W. Philips, D. Popescu, & P. Scheunders (Eds.), *Advanced Concepts for Intelligent Vision Systems* (pp. 542–553). Springer International Publishing.

Verburg, M., & Menkovski, V. (2019, May). Micro-expression detection in long videos using optical flow and recurrent neural networks. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)* (pp. 1–6). IEEE.

Wang, S.-J., He, Y., Li, J., & Fu, X. (2021). MESNet: A

convolutional neural network for spotting multi-scale micro-expression intervals in long videos. *IEEE Transactions on Image Processing*, *30*, 3956–3969. https://doi.org/10.1109/TIP.2021.3064258

Wang, S. J., Wu, S., & Fu, X. (2016, November). A main directional maximal difference analysis for spotting micro-expressions. In C.-S. Chen, J. Lu, & K.-K. Ma (Eds.), *Computer Vision – ACCV 2016 Workshops. ACCV 2016. Lecture Notes in Computer Science* (vol 10117, pp. 449–461). Springer, Cham.

Wen, J., Yang, W., Wang, L., Wei, W., Tan, S., & Wu, Y. (2020, December). Cross-database micro expression recognition based on apex frame optical flow and multi-head self-attention. In *International Symposium on Parallel Architectures, Algorithms and Programming* (pp. 128−139). Springer, Singapore.

Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., & Fu, X. (2014). CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PloS One*, *9*(1), e86041.

Yan, W.-J., Wu, Q., Liu, Y.-J., Wang, S.-J., & Fu, X. (2013, April). CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)* (pp. 1−7). IEEE.

Micro-expression spotting method based on human attention mechanism

LI Jingting[1], DONG Zizhao[1], LIU Ye[1,2], WANG Su-Jing[1,2], ZHUANG Dongzhe[3]

([1] *CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing* 100101, *China*)

([2] *Department of Psychology, University of Chinese Academy of Sciences, Beijing* 100049, *China*)

([3] *Public Security Behavioral Science Laboratory, People's Public Security University of China, Beijing* 100038, *China*)

Abstract: Micro-expressions are facial movements that are extremely short and not easily perceived, revealing the individual's hidden real emotions, and could be widely used in lies detection and other fields. The automatic research of micro-expression spotting is mainly limited by the small sample size. This project will address this problem by comprehensively using computer vision technology and cognitive psychology experimental methods. First, a behavioral-experimental paradigm combining eye-movement techniques and a presentation-judgment paradigm with subthreshold emotion priming effects was used to examine the cognitive mechanisms of selective attention allocation in micro-expression recognition and to refine the characteristic regions of interest in human recognition of micro-expressions. Second, based on the human attention mechanism, we propose a micro-expression spotting method based on a multi-branching self-supervised learning network, extracting structure-based, detail, spatio-temporal variation, and depth features of video samples. This research will achieve theoretical and technological breakthroughs in the field of automatic micro-expression spotting, and lay the foundation for the application of micro-expression spotting in realistic and complex scenarios.

Key words: micro-expression spotting, small sample problem, human attention mechanism, self-supervised learning, depth information